

## Grau en Estadística

---

**Títol:** Avaluació del tipus de matèria prima i d'antioxidant en la producció d'olis de l'empresa "Additius".

**Autor:** Gemma Garcia de la Fuente

**Director:** Isaac Subirana

**Departament:** Genètica, Microbiologia i Estadística

**Convocatòria:** 22 de Gener de 2020



UNIVERSITAT DE  
BARCELONA



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH

Facultat de Matemàtiques i Estadística

# Additius

Gemma Garcia de la Fuente

Isaac Subirana Cachinero

Grau d'estadística a la Universitat de Barcelona i la Universitat Politècnica de Catalunya

22 of gener, 2020

## Abstract

---

Català: L'objectiu principal del treball és determinar un mètode per seleccionar la matèria primera i millorar la estabilitat dels productes acabats d'una empresa agroquímica. Per fer-ho realitzarem dos dissenys d'experiments, cada un enfocat a resoldre una part del objectiu. Al llarg del treball veurem com s'ha identificat el tipus de disseny en cada cas, el temps que ha de durar la recollida de dades, la estimació dels recursos necessaris, controlant l'aleatorietat per evitar biaixos, supervisant la recollida de dades durant la durada de l'experiment i, per últim, realitzant les proves estadístiques necessàries per poder interpretar els resultats de l'experimentació. English: The main work objective is find a way to select the raw materials and improve the stability of the final products of a agrochemical company. To do it, we will perform two experiment designs oriented to solve each one of the goals aspects. Throughout the work we will see how the type of design has been identified in each case, the time required for data collection, estimate the necessary resources, controlling the randomness to avoid biases, monitoring data collection over the duration of the experiment and, finally, performing the statistical tests necessary to interpret the results of the experiment.

*Keywords:* Disseny d'experiment, Models lineals, Ois, Estabilitat producte

---

## Contents

<b>1</b>	<b>Introducció</b>	<b>3</b>
1.1	Paraules clau . . . . .	3
1.2	Classificació AMS . . . . .	3
1.3	Empresa . . . . .	4
<b>2</b>	<b>Metodologia</b>	<b>6</b>
2.1	Calendari . . . . .	6
2.2	Les dades . . . . .	7
2.3	Variables . . . . .	8
2.4	Maquinària . . . . .	10
2.5	Els lots . . . . .	12

<b>3</b>	<b>Exeperiments</b>	<b>13</b>
3.1	Olis . . . . .	13
3.1.1	Mostra . . . . .	13
3.1.2	Resultats . . . . .	15
3.1.2.1	Anàlisi descriptiu . . . . .	15
3.1.2.2	Anàlisis gràfic . . . . .	16
3.1.3	Inferència . . . . .	20
3.1.4	Modelització . . . . .	22
3.1.4.1	Selecció model . . . . .	25
3.1.4.2	Predicció model . . . . .	27
3.2	Antioxidants . . . . .	29
3.2.1	Mostra . . . . .	29
3.2.2	Resultats . . . . .	32
3.2.2.1	Anàlisis descriptives . . . . .	32
3.2.2.2	Anàlisis gràfiques . . . . .	32
3.2.3	Modelització . . . . .	33
<b>4</b>	<b>Conclusions</b>	<b>38</b>
<b>5</b>	<b>Limitacions</b>	<b>39</b>
<b>6</b>	<b>Recomanacions per futures investigacions</b>	<b>39</b>
<b>7</b>	<b>Bibliografia</b>	<b>40</b>

# 1 Introducció

Aquest escrit és la memòria del projecte realitzat per a una empresa dedicada a la producció d'agroquímica. L'empresa es dedica a la creació d'un colorant natural, aquest pigment es barreja amb l'aliment de pollastres i gallines per millorar el color de la pell dels animals i del rovell dels ous.

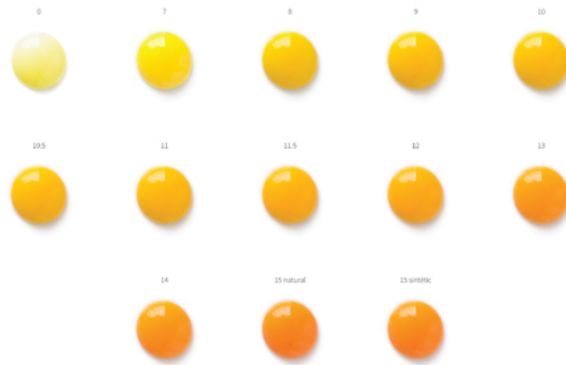


Figure 1: Ous



Figure 2: Pollastre

L'objectiu d'aquest treball és optimitzar el rendiment de la producció i estabilitzar el producte de l'empresa.

## 1.1 Paraules clau

Anàlisi estadística, modelització, anàlisi olis, comparació múltiple, Disseny d'experiment, Models lineals, Olis, Estabilitat producte, etc

## 1.2 Classificació AMS

La temàtica d'aquest escrit segons la Societat Matemàtica Americana AMS correspon a la secció d'Informàtica (68-XX).



### 1.3 Empresa

La producció dels pigments es fa a partir de l'oli extret d'una planta que s'anomena Marigold (clavell moro), aquest oli es processa i s'absorbeix en un silicat per transformar-ho en pols convertir-se en el producte final que ofereixen als clients.



Figure 3: Producte inicial i producte final

Per assolir l'objectiu s'estableixen dos projectes:

- Dissenyar un model per identificar les característiques dels olis que s'oxiden menys
- Crear un disseny per identificar el millor antioxidant pel seu producte

*Definim oxidació com:* L'oxidació com a tal es refereix al procés químic que implica la pèrdua d'electrons per part d'una molècula, àtom o ió. Quan això ocorre, diem que la substància ha augmentat el seu estat d'oxidació. En el cas que ens ocupa, l'oxidació disminueix la capacitat pigmentant (contingut del component actiu) del producte.

Quan arriben els olis, es fan ús seguit de proves per determinar-ne el rendiment en xantofil·les (principi actiu), i la resta de característiques físico-químiques. Un cop es té aquesta informació prèvia, els olis passen a producció. Després d'aquesta se n'obté el rendiment fent un balanç de matèria activa:  $(\text{kg} \times \text{concentració}) \text{ finals} / (\text{kg} \times \text{concentració}) \text{ inicials}$ . El següent és l'absorció (Passar de líquid a sòlid), després de l'absorció també es pot calcular el rendiment de la mateixa tot i que és menys significatiu. Finalment cal estandaritzar, afegint material inert per obtenir la concentració desitjada, per obtenir el producte acabat.

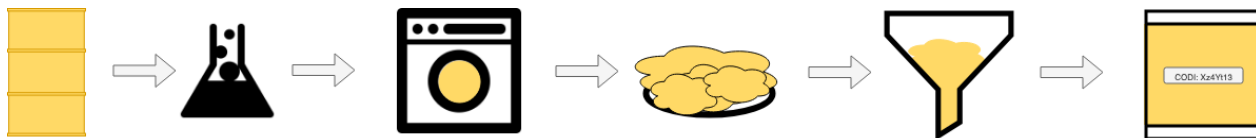


Figure 4: Producció

Per avaluar la qualitat i obtenir un antioxidant es proposen dos experiments, per a cada un és dura a terme des del disseny de l'experiment fins a l'anàlisi del resultat, durant el procés es controlaran: la recollida de dades, l'aleatorietat, els biaixos, l'avaluació dels possibles problemes, ...

Els dos experiments segueixen el mateix camí que fan els olis en la producció normal de l'empresa fins que s'acaba l'absorció. S'agafaran les mostres necessàries per calcular el rendiment de la producció i, posteriorment, establir diferents moments temporals, al principi més propers i després més espaiats, on farem un control del rendiment per determinar l'efecte de l'oxidació. Aquestes mostres es conservaran en una càmera climàtica en condicions constant d'humitat i temperatura.

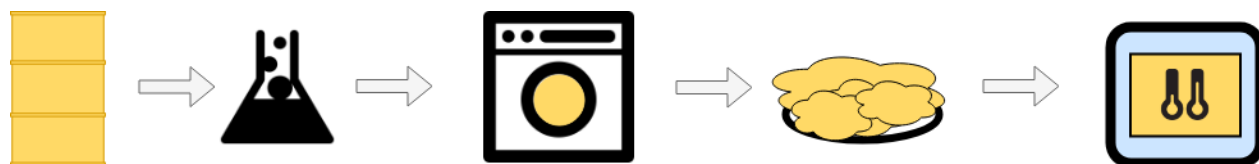


Figure 5: Producció

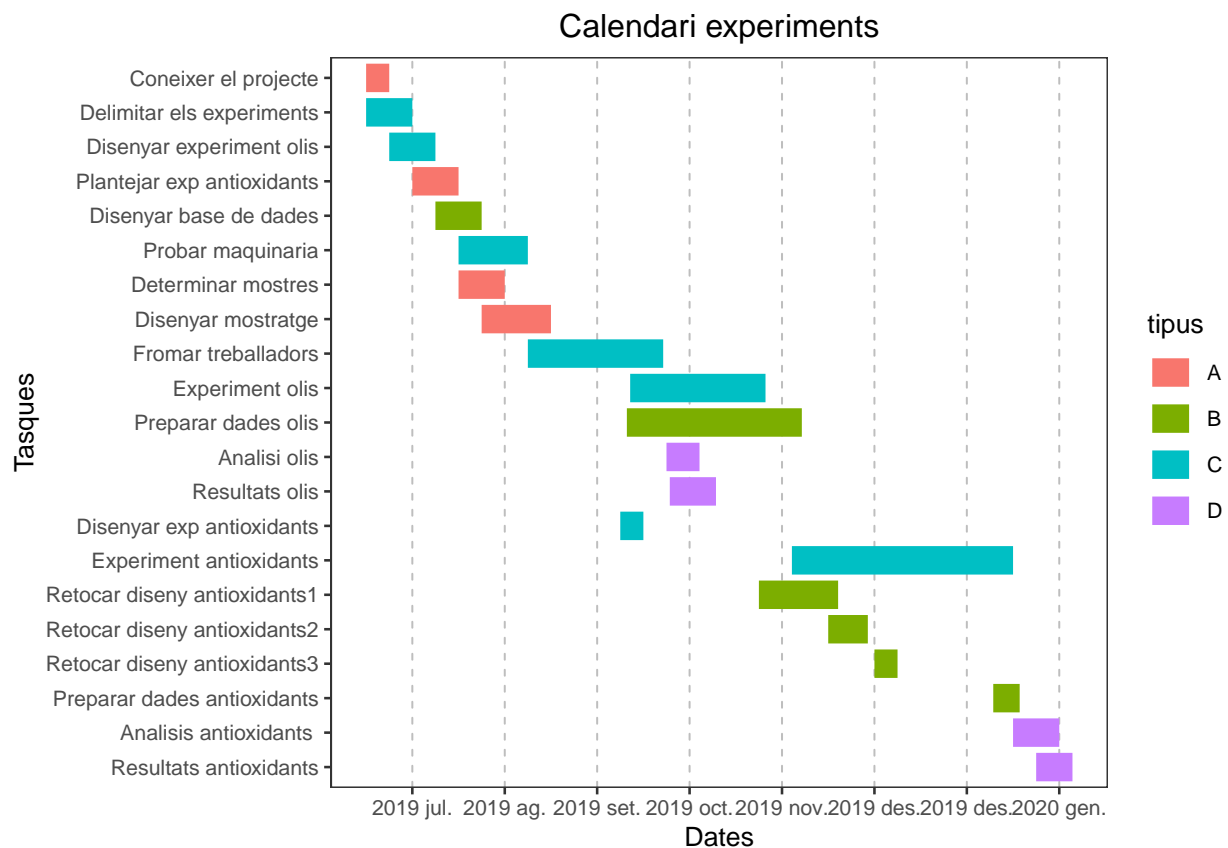
## 2 Metodologia

### 2.1 Calendari

A l'inici del projecte s'estableix un calendari que servirà de guia de cara a la planificació de les diferents tasques. Destaquem 4 categories en funció del tipus de tasca:

- A: Tasques de redacció
- B: Tasques de processament de dades
- C: Tasques dels experiments
- D: Tasques d'anàlisi

A continuació podem veure un diagrama de Gantt on s'especifica el calendari provisional previst per a la realització dels experiments, podem veure la durada de cada tasca i l'ordre amb què s'aniran realitzant:



## 2.2 Les dades

L'empresa posa a disposició tot el material necessari així com la maquinària i els treballadors. Es disposa de 10 lots de diferents olis, és a dir diferents orígens i dates de producció. De cada lot s'extreuen 3 mostres, per tant tindrem un total de 30 observacions.

Num	ID lot	Lot
1	91016298-1	Proveidor 1
2	91016311-2	Proveidor 1
3	91016301-1	Proveidor 1
4	91016407-1	Proveidor 2
5	91016472-1	Proveidor 1
6	91016428-1	Proveidor 1
7	91016487-1	Proveidor 4
8	91016487-2	Proveidor 3
9	91016498-2	Proveidor 1
10	91016429-1	CProveidor 5

El que pretenem és identificar els olis que donen millor rendiment en la producció i relacionar-los amb les característiques analítiques. D'aquesta manera l'empresa podrà fer una selecció de matèries primeres, a partir de les mostres representatives dels lots que els ofereixen. Aquest és la primera part de l'experiment. El que volem d'aquests olis són dos elements:

- El seu rendiment. Valor escalar obtingut directament després de realitzar la reacció. Aquesta és la mesura important per l'empresa, ja que una millora en el rendiment suposa una reducció en el cost de la producció.
- La seva estabilitat: Sèrie temporal, un cop realitzada la reacció es conserven els olis homogeneïtzant les condicions ambientals per tant d'avaluar com evoluciona el producte, si es fa malbé ràpid o lent.

Un cop obtinguts aquests valors es vol estudiar si els olis bons es caracteritzen per uns valors concrets de les variables recollides de cada lot. Si no es trobés cap perfil característic dels olis, en tenir-ne una quantitat addicional, se'ls podrien avaluar altres variables a posteriori.

Un cop realitzat el primer experiment triarem els que donin millor i pitjor rendiment per utilitzar-los al segon experiment.

El segon experiment pretén ralentir el procés d'oxidació en el producte acabat. Per fer-ho cal identificar el millor antioxidant. Fins ara l'empresa treballava amb un antioxidant (producte per posposar l'oxidació), que els funcionava molt bé però que ha estat retirat del mercat. Com que no existeix cap anàleg conegut igual de bo, en l'experiment el que es vol és comparar diferents antioxidants per trobar quin és més adient.

L'empresa disposa de 2 antioxidants per avaluar: BHT, BHA i AG (Àcid Gàlic)

El que farem és afegir diferents antioxidants a les mostres de cada lot i, després de la producció, guardar mostres del producte acabat en condicions estables d'humitat i temperatura per determinar quin és el més efectiu.

El procediment que segueixen les mostres en ambdós experiments és equivalent, l'única diferència és que a l'inici del procés se li afegeix o no l'antioxidant.

## 2.3 Variables

Així doncs, les dades es recullen 'in situ' es a dir, es fan els analisi a la mateixa empresa, excepte en alguns casos concrets en s'envien a laboratoris externs.

- Lot: identificador
- Oleo: origen
- Densitat
- Luteina
- Granulometria tamís 1.5 i 1 mm
- L,a,b : paràmetres identificadors de color
- Humitat en estufa a 135°
- Rendiment
- Zeaxantina
- Epóxids
- pH en moment inicial i al cap de 7, 14, 21, 28, 35 i 42 dies

Per a cada bidó d'oli s'obtéindrà una mostra de 500 g que s'enviarà directament al laboratori. Una part d'aquestes mostres seran enviades a laboratoris externs. En acabar el procés es guardarà el producte en pols (40kg aprox) per al control de l'estabilitat. Per aquest es guardarà una petita mostra en la cambra climàtica la qual està programada amb 40°C i 75% d'humitat relativa). D'aquesta mostra es comprovarà el rendiment als 7, 14, 21, 28, 35 i 42 dies.

L'estructura que tindrà la base de dades es caracteritza amb el següent diagrama d'entitat relació:

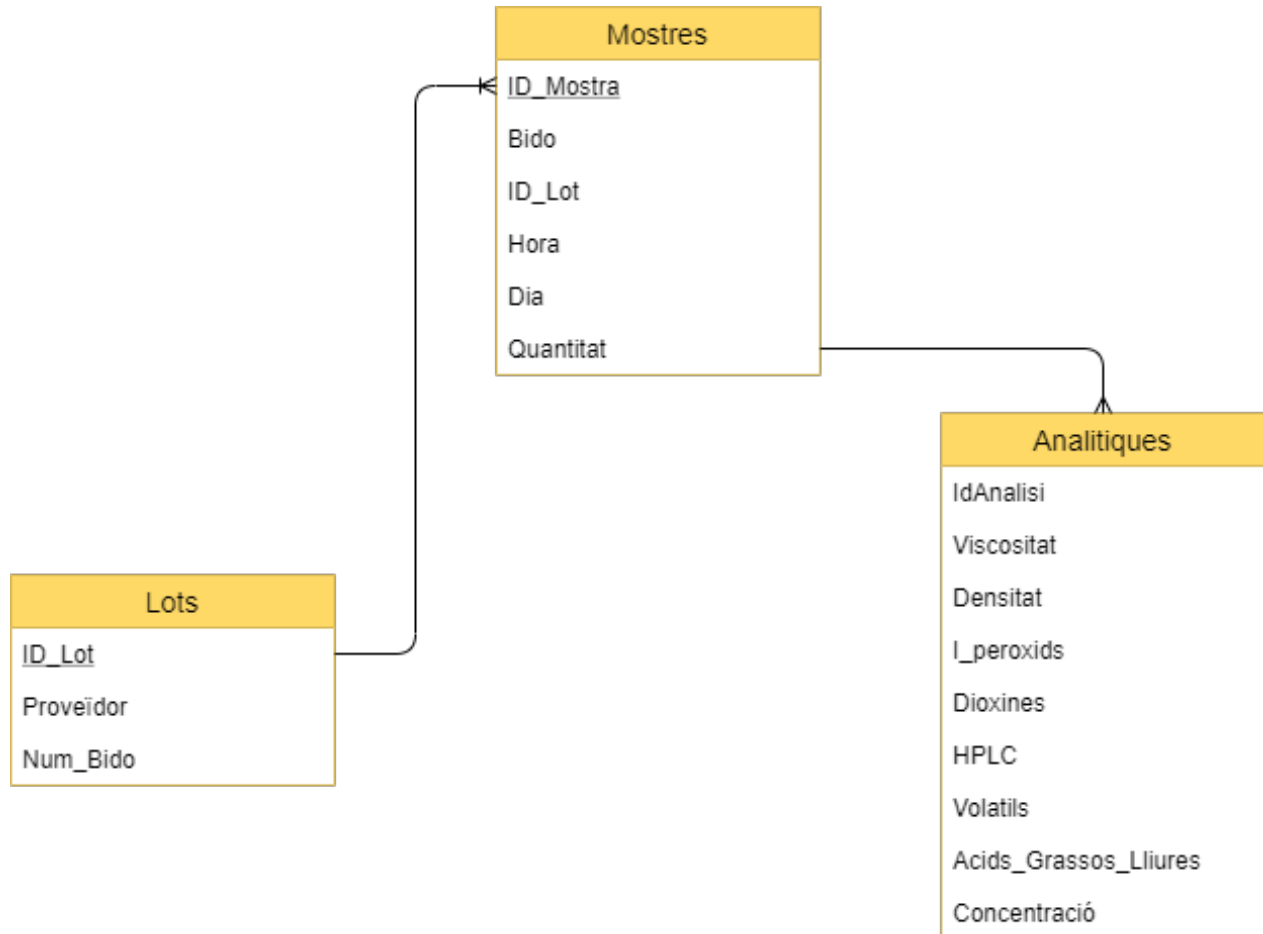


Figure 6: Diagrama Entitat Relació

## 2.4 Maquinària

Per a dur a terme tot el procés així com el recull de la informació l'empresa disposa d'una planta pilot on realitzar les reaccions necessàries a les quals sotmetre el producte i un laboratori de recerca per realitzar l'experimentació de forma controlada.

Per a la planta pilot fou necessari dissenyar un reactor a petita escala, i una barrejadora adient, tot seguit en podem veure un esquema i una foto.

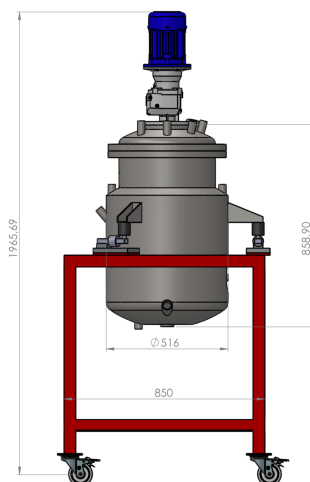


Figure 7: Dimencions Reactor

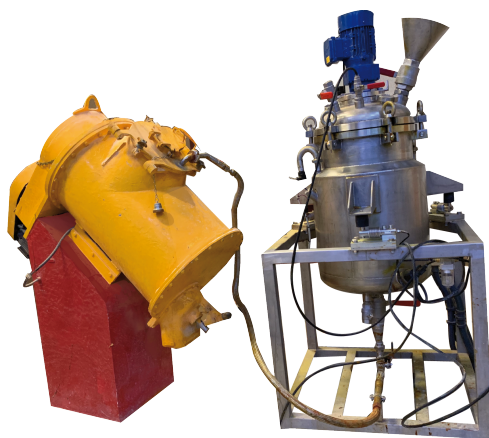


Figure 8: Reactor

El funcionament del reactor aquest està sincronitzat a un panell de control al qual s'hi pot accedir des de l'ordinador del laboratori, aquest panell és interactiu i per tant permet fixar paràmetres com per exemple: la velocitat amb la qual es remena l'oli, la temperatura, etc.

Per mantenir les condicions atmosfèriques estables i per tant despreocupar-se de variables ambientals que podrien afectar molt al producte, s'ha adquirit una cambra climàtica:



Figure 9: Cambra climàtica

En l'interior te uns prestatges per dipositar les provetes, i al seu torn també té una pantalla en la qual es poden ajustar paràmetres relacionats amb les condicions atmosfèriques que es volen mantenir, aquestes han de ser constats durant tot el projecte per tal de poder suprimir la variabilitat provinent de la temperatura, humitat, ...



## 2.5 Els lots

És molt important predefinir les característiques de les mostres per a tal d'homogeneïtzar-ne les condicions entre olis i reduir biaixos. Per aquest motiu l'empresa ha establert un protocol.

Com ja s'ha comentat disposem de 10 lots, cada lot té 4 bidons de 200kg. De cada lot s'escull un bidó i es posa en una cambra a 65° durant 48h, s'agitarà el producte per homogeniar-lo. De cada bidó es trauran 6 mostres de 10 kg que es guardaran en sacs d'alumini soldats.

No s'afegirà cap producte additiu a l'oli. Aquests sacs seran els que s'utilitzaran per a les reaccions.



Figure 10: Sacs

El producte que sobri dels bidons s'utilitzarà en la producció normal de l'empresa, recordem que per a cada lot encara disposem d'altres 3 bidons per poder realitzar més proves posteriorment.

A continuació tenim una taula amb l'entrada dels bidons a la cambra:

Entrada	Salida	Interno	Proveedor
16-09-19	18-09-19	6298-1	Proveidor 1
16-09-19	18-09-19	6301-1	Proveidor 1
16-09-19	18-09-19	6311-2	Proveidor 1
16-09-19	18-09-19	6407-1	Proveidor 2
17-09-19	19-09-19	6428-1	Proveidor 2
17-09-19	19-09-19	6429-1	Proveidor 5
17-09-19	19-09-19	6472-1	Proveidor 1
17-09-19	19-09-19	6487-1	Proveidor 4
18-09-19	20-09-19	6487-2	Proveidor 3
18-09-19	20-09-19	6498-2	Proveidor 1

### 3 Exeperiments

Són dos els experiments que es presenten en aquest escrit, el primer és el dels olis i el segon el dels antioxidants. El que fa la empresa és :

Treballar amb uns olis que provenen de la planta Marigold. Aquests olis són la materia primera dels pigments naturals que ofereixen als seus clients.



Figure 11: Marigold

#### 3.1 Olis

Han observat al llarg del temps que alguns olis s'oxiden més de pressa que altres. L'objectiu és identificar les característiques dels olis que s'oxiden més lentament. Es proposa doncs fer un experiment en el qual se sotmeten els olis a una bateria de proves pre i post producció i, a l'hora, s'avalua la seva concentració (Xantrofila/Kg), d'on obtenim el rendiment, pre i post per veure l'oxidació de cada oli en concret. Aquests olis se sotmeten a un procés de producció que converteix els olis en pols. Les proves que es realitzen permeten calcular la pèrdua de cada un dels olis. Això permetrà identificar els olis més estables i, amb l'anàlisi estadístic, mirarem de determinar les característiques que els identifiquen. Cal tenir en compte que el procés de producció sempre genera una oxidació que està entre el 1% i el 10%.

##### 3.1.1 Mostra

En aquest primer experiment com que volem conèixer la qualitat pròpia de l'oli, no s'afegeix cap producte protector ni antioxidant. Per tant es treballa amb els sacs descrits anteriorment com a mostres inicials en el procés de producció.

Per tal de preservar unes òptimes condicions s'ha fet un llistat aleatori amb Excel per establir l'ordre amb què es faran les reaccions de les diferents mostres. En la taula que es presenta a continuació és veure l'ordre en què es realitzen les reaccions, el sac i el lot que pertany cada una.

Ordre	Sac	Lot
1	6298-1	Proveidor 1
2	6487-1	Proveidor 4
3	6487-1	Proveidor 4
4	6487-2	Proveidor 3
5	6498-2	Proveidor 1
6	6472-1	Proveidor 1
7	6428-1	Proveidor 2
8	6301-1	Proveidor 1
9	6429-1	Proveidor 5
10	6311-2	Proveidor 1
11	6301-1	Proveidor 1
12	6428-1	Proveidor 2
13	6429-1	Proveidor 5
14	6407-1	Proveidor 2
15	6428-1	Proveidor 2
16	6498-2	Proveidor 1
17	6487-2	Proveidor 3
18	6498-2	Proveidor 1
19	6407-1	Proveidor 2
20	6429-1	Proveidor 5
21	6487-2	Proveidor 3
22	6407-1	Proveidor 2
23	6301-1	Proveidor 1
24	6298-1	Proveidor 1
25	6298-1	Proveidor 1
26	6472-1	Proveidor 1
27	6311-2	Proveidor 1
28	6487-1	Proveidor 4
29	6472-1	Proveidor 1
30	6311-2	Proveidor 1

El **procediment** que es realitza per a cada oli consisteix en:

1. Carregar el reactor amb el sac que toca.
2. Agitar el producte
3. Fer la reacció
4. Recollir la mostra i analitzar-ne el rendiment

### 3.1.2 Resultats

Els resultats que s'han obtingut els analitzem a continuació, realitzarem unes anàlisis descriptives numèriques i gràfiques i posteriorment ajustarem un model per veure quines variables explicatives caracteritzen millor els olis.

#### 3.1.2.1 Anàlisi descriptiu

rendiment	pH_inicial	Densidad	Gran1.5	Gran1
Min. :0.812	Min. :10.1	Min. :0.582	Min. :97.1	Min. :94.7
1st Qu.:0.896	1st Qu.:10.3	1st Qu.:0.591	1st Qu.:99.4	1st Qu.:96.9
Median :0.939	Median :10.4	Median :0.600	Median :99.7	Median :98.2
Mean :0.930	Mean :10.4	Mean :0.598	Mean :99.5	Mean :97.7
3rd Qu.:0.965	3rd Qu.:10.5	3rd Qu.:0.602	3rd Qu.:99.8	3rd Qu.:98.7
Max. :1.012	Max. :11.0	Max. :0.618	Max. :99.9	Max. :99.0

L	a	b	Hum	pH_07
Min. :54.5	Min. :21.3	Min. :63.0	Min. :21.8	Min. :10.1
1st Qu.:58.6	1st Qu.:23.2	1st Qu.:67.1	1st Qu.:22.7	1st Qu.:10.3
Median :60.0	Median :24.4	Median :68.7	Median :23.1	Median :10.4
Mean :59.5	Mean :24.2	Mean :68.2	Mean :23.4	Mean :10.4
3rd Qu.:60.6	3rd Qu.:25.2	3rd Qu.:69.2	3rd Qu.:24.0	3rd Qu.:10.6
Max. :61.6	Max. :27.0	Max. :71.3	Max. :26.3	Max. :10.8

pH_14	pH_21	pH_28	pH_35	pH_42
Min. :10.2	Min. :10.4	Min. :10.4	Min. :10.4	Min. :10.5
1st Qu.:10.4	1st Qu.:10.5	1st Qu.:10.6	1st Qu.:10.6	1st Qu.:10.6
Median :10.6	Median :10.6	Median :10.6	Median :10.6	Median :10.7
Mean :10.6	Mean :10.7	Mean :10.7	Mean :10.7	Mean :10.7
3rd Qu.:10.8	3rd Qu.:10.9	3rd Qu.:10.7	3rd Qu.:10.7	3rd Qu.:10.7
Max. :10.9	Max. :11.2	Max. :11.0	Max. :11.0	Max. :11.0

Luteina	Zeaxantina
Min. :80.4	Min. :4.32
1st Qu.:84.2	1st Qu.:4.89
Median :86.1	Median :5.14
Mean :85.5	Mean :5.34
3rd Qu.:86.5	3rd Qu.:5.44
Max. :87.8	Max. :9.31

### 3.1.2.2 Anàlisi gràfic

Com que treballem amb dades numèriques realitzem un boxplot per a cada una de les variables, a més a més mirem si les variables tenen valors extrems. El criteri que utilitzem per identificar un valor extrem com outlier és a través dels boxplots, la mateixa funció `boxplot()` et mostra els valors que considera fora dels límits:

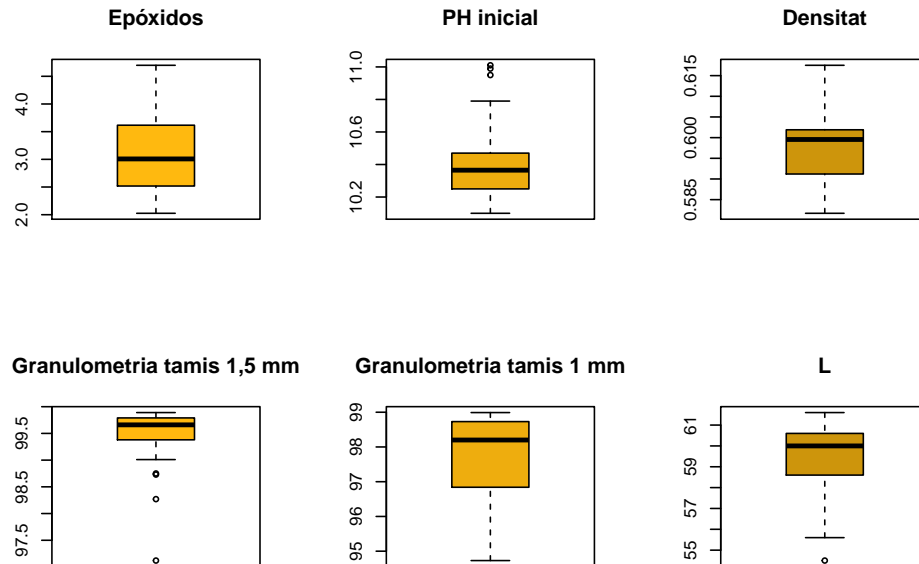


Figure 12: Variables

- **Epóxidos**, no trobem cap valor mancant.
- **L**, en aquesta podem observar un valor atípic. Al ser un paràmetre del color el fet de ser considerat valor atípic significa que el color del lot varia respecte a la majoria. El valor crític pren una puntuació de 54.5.
- **PH inicial**, els valors que veiem a continuació són els considerats atípics en una primera anàlisi visual, caldrà tenir en compte els lots per si posteriorment tenen un comportament anormal. Les puntuacions que registren aquets lots són: 10.99, 11.01 i 10.95.
- **Granulometria tamis 1,5 mm**, en aquesta variable també s'identifiquen de forma visual alguns valors atípics: 98.73, 98.75, 98.27 i 97.12.
- **Granulometria tamis 1 mm**, en aquesta podem observar un valor atípic
- **Densitat**, no té valors atípics.

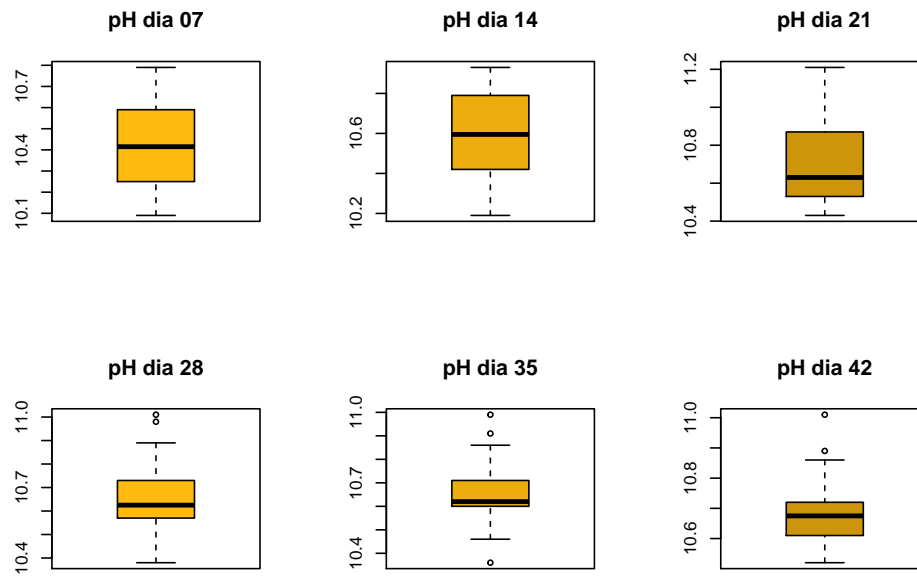


Figure 13: Variables

- Ph al cap de 7 dies, 14 dies, 21 dies , en aquestes no trobem cap valor anormal.
- Ph al cap de 28 dies, sí que trobem valors atípic: 10.98 11.01
- Ph al cap de 35 dies, trobem 3 valors atípics: 10.99, 10.91 i 10.36.
- Ph al cap de 42 dies, trobem 2 valors atípics: 10.89 i 11.01

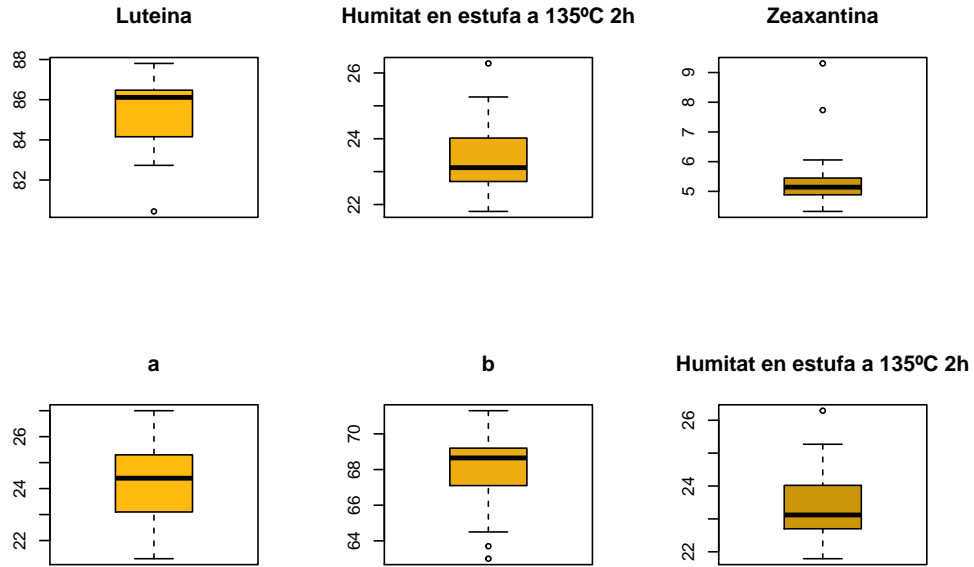


Figure 14: Variables

- **b**, en aquesta podem observar dos valors atípics: : 63.7 63.0 . Ens torbem en la mateixa situació del cas anterior.
- **a**, En aquest paràmetre de color, no trobem cap valor atípic.
- **Humitat en estufa a 135°**, en aquesta podem observar un valor atípic amb valor de 26.29.
- **Luteina**, si que trobem un valor atípic que arriba a 80.44.
- **Zeaxantina**, trobem dos valors atípics: 7.736 i 9.307.

La **variable resposta** és la variable Rendiment, que és la que recull la qualitat de l'oli, fem una petita descripció:

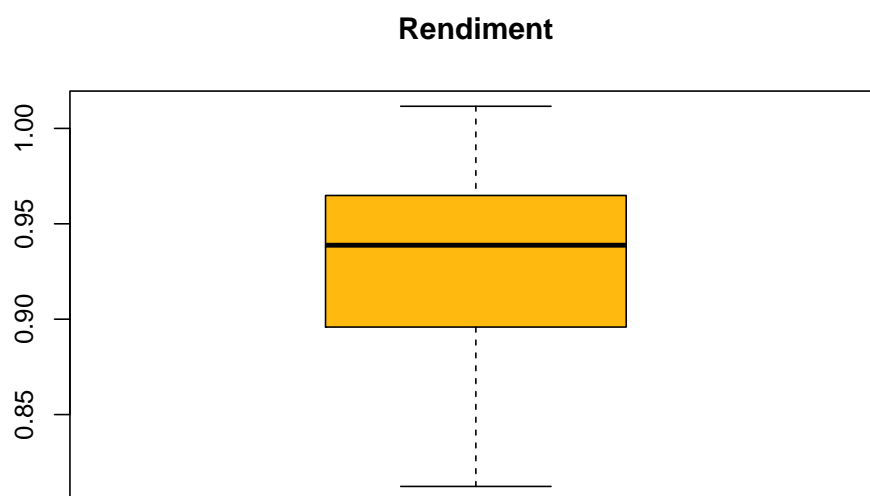


Figure 15: Variable resposta

Fem un test de normalitat i un test de homocsedasticitat, a més a més de per estudiar el comportament de la variable ens serveix per comprovar si es satisfent els supòsits pels tests estadístics:

**Test normalitat:** farem servir un test de Shapiro - Wilks, el contrast per aquest test és:

$$H0 : \text{Normalitat}$$

$$H1 : \text{No normalitat}$$

Obtenim un estadistic de  $w = 0.9$  al qual se li associa un p-valor molt elevat: 0.27, per tant direm que les dades de la variable rendiment segueixen una distribució normal. Veiem un gràfic Q-Qplot a continuació que també ens ho mostra:



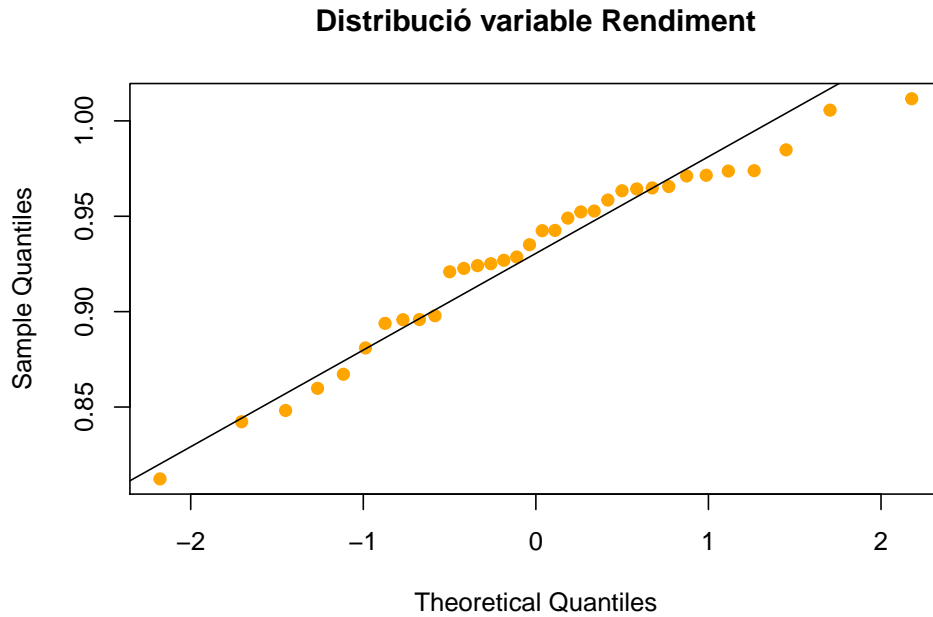


Figure 16: Normalitat variable resposta

En excepció d'alguns valors que es desvien de la recta, sobretot en els extrems, veiem que efectivament arribaríem a la mateixa conclusió que el test estadístic: no tenim motius per rebutjar la hipòtesi nul·la.

### 3.1.3 Inferència

Tenim una variable amb el registre dels diferents **Orígens dels olis**, per aquest motiu fem una comparació de mitjanes, tenim una variable quantitativa (rendiment) i una variable on es recull els orígens dels olis, volem estudiar si hi ha diferències entre el rendiment i l'origen de la matèria primera. En primer lloc mostrem una descriptiva amb les dades agrupades:

Oleo	mean
Proveïdor 1	0.9286
Proveïdor 2	0.9321
Proveïdor 3	0.9318
Proveïdor 4	0.9634
Proveïdor 5	0.9221

Veiem que les diferències en el rendiment per a cada grup d'origen són molt subtils, fem una representació gràfica:

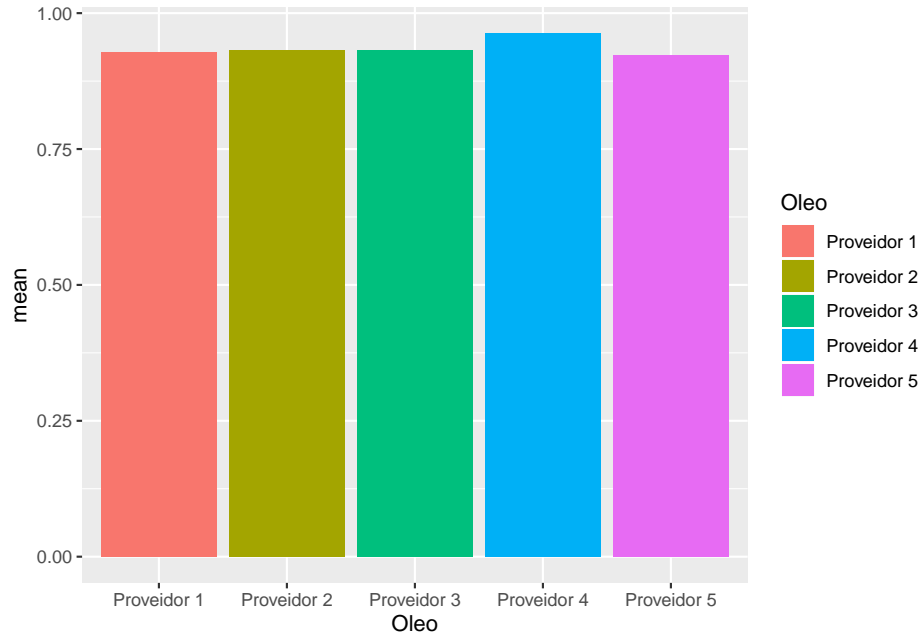


Figure 17: Normalitat variable resposta

Per a fer la comparació entre els grups, com que ja hem vist que les dades són normals, apliquem un test ANOVA:

Creem el model per contrastar la significació del factor nivell Oleo en el rendiment tenim el següent contrast d'hipòtesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$H_1 : \text{Com a mínim un parell de mitjanes sn diferents}$$

Table 9: Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>dd\$Oleo</b>	4	0.001376	0.000344	0.1368	0.9674
<b>Residuals</b>	29	0.07294	0.002515		

Amb el test obtenim un p-valor molt elevat, això ens porta a no tenir motius per rebutjar la hipòtesi nul·la. Per tant, considerem que l'origen dels olis no té un efecte estadísticament significatiu sobre el rendiment. Amb els resultats del test el que fem és considerar que l'origen de les variables no té un efecte estadísticament significatiu i per tant no la tindrem en compte en la modelització (Badiella, L 2011).

### 3.1.4 Modelització

Ajustem un primer model on la variable important és el rendiment i les variables explicatives són totes les que disposem. El que farem és ajustar una regressió lineal per veure quines variables explicatives ens ajuden a preveure el rendiment dels olis (Seber, G. A., & Lee, A. J. 2012).

Per assegurar-nos que el model és significatiu fem un test Òmnibus:

$H_0$  : *El model ajusta b les dades*

$H_1$  : *El model no ajusta b les dades*

Per fer el test necessitem ajustar també el model nul. Ajustem un model lineal generalitzat de família gaussiana, on la funció enllaç és 1.

Fem un test ANOVA comparant els dos models per poder obtenir l'estadístic F així com el seu p-valor associat per decidir si el model es considera significatiu comparat amb un model nul.

Table 10: Analysis of Deviance Table

Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)
16	0.02015				
33	0.07432	-17	-0.05417	2.531	0.03498

El model és significatiu, és a dir: les variables explicatives expliquen les dades. El p-valor que obtenim és inferior al nivell de significació (5%).

El que fem és utilitzar la funció `step()` de R per veure quines variables són més significatives. Aquesta funció el que fa és executar un procés iteratiu. El procés inicia amb totes les variables i va descartant una a una en funció del si es redueix el AIC.

Table 11: Analysis of Deviance Table

Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)
26	0.02501				
33	0.07432	-7	-0.04931	7.324	0.00006984

Veiem que el model amb les dades seleccionades per la funció `step` surt significatiu, és millor que el model nul. Però també és interessant comparar-lo amb el model saturat (és a dir el que conté totes les variables) amb el model seleccionat amb la funció `step()`:

Table 12: Analysis of Deviance Table

Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)
26	0.02501				
16	0.02015	10	0.004861	0.386	0.9347

Veiem que el model simplificat es considera estadísticament igual al model saturat. Així i tot hem vist que el simplificat respecte al nul aporta informació mentre que el saturat no. Veiem a continuació l'aportació de cada variable en la predicció:

Table 13: Analysis of Deviance Table (Type II tests)

	LR Chisq	Df	Pr(>Chisq)
<b>pH_inicial</b>	3.711	1	0.05406
<b>Gran1.5</b>	6.912	1	0.008561
<b>pH_14</b>	2.36	1	0.1245
<b>pH_21</b>	12.51	1	0.0004049
<b>Luteina</b>	7.338	1	0.006751
<b>Zeaxantina</b>	7.82	1	0.005168
<b>Epoxidos</b>	2.038	1	0.1534

Amb els resultats obtinguts veiem les variables que realment aporten informació per a predir el rendiment, creem un nou model sense les variables que no han sortit significatives: : pH\_14, i Epoxidos amb p-valors molt elevats per tant les traïem.

Table 14: Analysis of Deviance Table (Type II tests)

	LR Chisq	Df	Pr(>Chisq)
<b>pH_inicial</b>	9.132	1	0.002512
<b>Gran1.5</b>	10.36	1	0.001291
<b>pH_21</b>	8.23	1	0.00412
<b>Luteina</b>	7.225	1	0.00719
<b>Zeaxantina</b>	4.303	1	0.03804

Totes les variables ens surten significatives.

Amb la simplificació aplicada al model i l'anàlisi de variables importants el que fem és comparar aquest model amb el primer model ajustat a partir del criteri AIC:

Table 15: Analysis of Deviance Table

Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)
26	0.02501				
28	0.02946	-2	-0.004449	2.313	0.119

Tot i tenir un menor nombre de paràmetres el model sembla que no millora.

El que fem ara és utilitzar la funció `step()` però especificant que volem utilitzar el criteri BIC en comptes del AIC. Perquè la funció canviï de criteri afegim un paràmetre:  $k = \log(n)$ .

Table 16: Analysis of Deviance Table

Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)
28	0.02946				
33	0.07432	-5	-0.04486	8.529	0.00005292

Igual que passa amb el model generat amb el criteri AIC el model generat amb el criteri BIC és millor que el model nul, ara comparem el model amb el model saturat:

Table 17: Analysis of Deviance Table

Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)
28	0.02946				
16	0.02015	12	0.00931	0.6161	0.7998

Respecte al model saturat no millora, és a dir: considerem que els dos models són iguals de bons per predir les dades. Analitzem l'aportació de cada variable respecte la variable resposta:

Table 18: Analysis of Deviance Table (Type II tests)

	LR Chisq	Df	Pr(>Chisq)
<b>pH_inicial</b>	9.132	1	0.002512
<b>Gran1.5</b>	10.36	1	0.001291
<b>pH_21</b>	8.23	1	0.00412
<b>Luteina</b>	7.225	1	0.00719
<b>Zeaxantina</b>	4.303	1	0.03804

Amb un nivell de significació de 5% ens quedem amb totes les variables, ja que els p-valors són petits. A més a més ens fixem que el model seleccionat amb el criteri BIC coincideix amb el model creat a partir de la selecció de variables significatives del model AIC.

### 3.1.4.1 Selecció model

Arribats a aquest punt fem una petita taula comparant els models, per fer-ho posarem la següent informació: nombre de paràmetres, AIC, BIC i R2 ajustat.

Models	Num_Param	AIC	BIC	R2_Ajustat
Model null	0	-107.8	-104.7	0
Model complet	17	-118.2	-89.17	0.4409
Model AIC 1	8	-130.8	-117.1	0.5558
Model AIC 2	5	-129.3	-118.6	0.5156
Model BIC 1	5	-129.3	-118.6	0.5329

Cal recordar que el que ens interessa són els valors mínims en AIC i BIC, i per altra banda el major en R2 Ajustat. Segons el criteri de AIC el millor model és el AIC 1, aquest mateix model té un BIC inferior als del model nul i complet, però superior a la resta de models simplificats. El BIC més elevat és el mateix pels models AIC 2 i BIC 1. Fem una taula simplificada per mostrar els 2 models preseleccionats com els millors AIC 1 i BIC 1:

Models	Num_Param	AIC	BIC	R2_Ajustat
Model AIC 1	8	-130.8	-117.1	0.5558
Model BIC 1	5	-129.3	-118.6	0.5329

Ambdós models són molt proper, ens els AIC i BIC, difereixen en 1.6 i 1.5 respectivament. Per tant per seleccionar el millor model utilitzem el  $R^2$  ajustat i amb aquest estadístic seleccionem el Model AIC 1 com a model que millor explica les dades. La R 2 ajustada indica el percentatge de variació explicat només per les variables independents que afecten realment la variable dependent, en el cas del model seleccionat és un 56%.

A continuació avaluem els residus el model seleccionat:

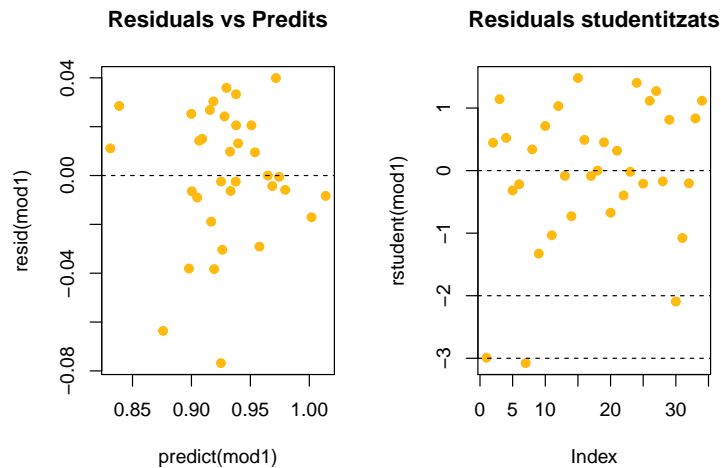


Figure 18: Ajust Model Inicial

Els residus semblen bastant normals, és a dir: la majoria es mantenen dins d'un interval de normalitat entre 2 i -2.

Avaluem si compleix la funció de variància, fem una representació gràfica:

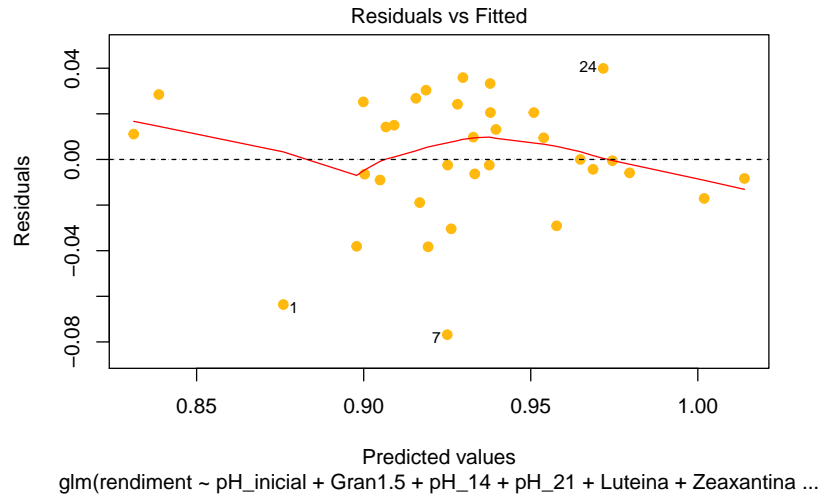


Figure 19: Estimació Segon Inicial

Veiem que alguns dels valors estan indicats com a valors atípics, així i tot en termes generals els residus són bastant bons: es compleix la linealitat i també la homocedasticitat.

Per acabar de validar el model fem un test de normalitat pels residus. El contrast que es planteja aquí és:

$$H_0 : \epsilon \sim N(0, 1)$$

$$H_1 : \epsilon \not\sim N(0, 1)$$

Table 21: Shapiro-Wilk normality test: `mod1$residuals`

Test statistic	P value
0.9356	0.0456 *

En aquest contrast obtenim un estadístic de 0.9398 amb un p-valor de 0.045. El valor del p-valor és molt proper al nostre nivell de significació de 0.05, en aquests casos normalment s'acceptaria que les dades són normals. Per acabar d'assegurar que realment fem bé en acceptar normalitat dels residus en fem un gràfic Q-Qplot :

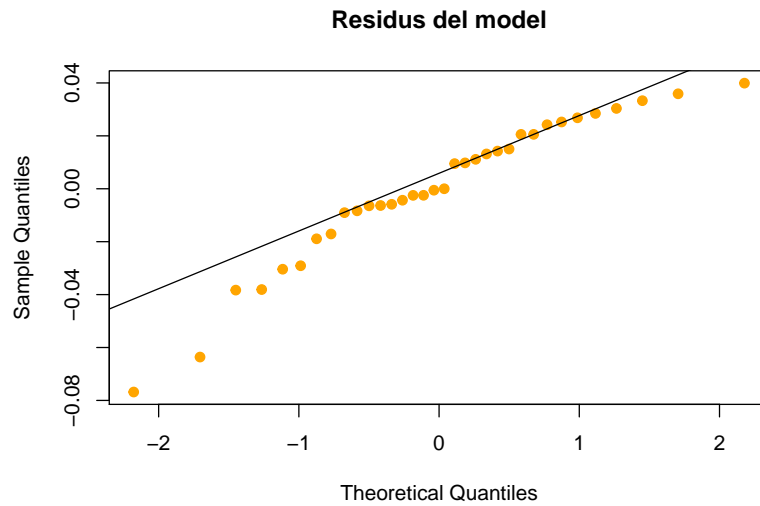


Figure 20: Residus del model

Veiem en el gràfic que els residus tenen un comportament anormal sobretot en l'extrem esquerre. Així i tot donem els residus per normals i per tant assumim que el model és útil per predir les dades.

### 3.1.4.2 Predicció model

Donem un cop d'ull a les prediccions del model:

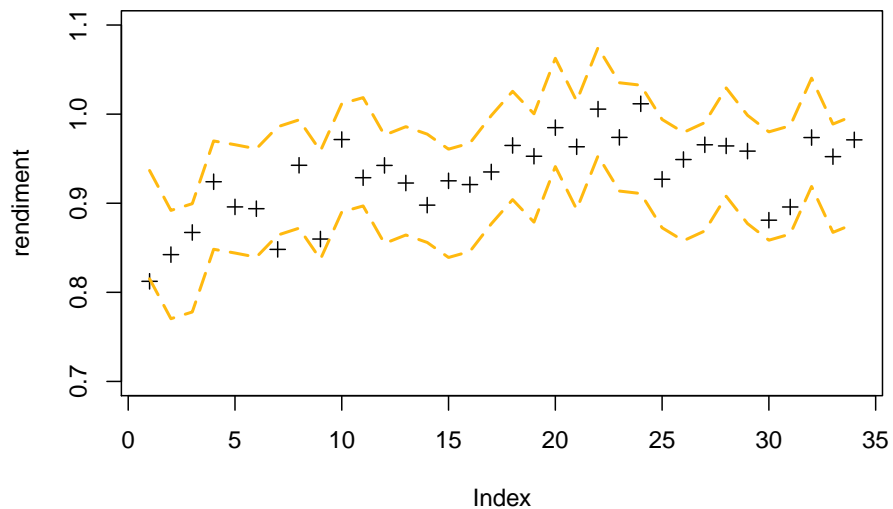


Figure 21: Predicció Model Inicial



Semblaria que el model ajusta bé. En el gràfic podem veure les prediccions en X i les línies grogues ens maquen l'interval de confiança. Veiem que pràcticament totes les observacions queden dins de l'interval i per tant que el model ajusta bé.

Amb això arribarem a la conclusió que les variables que millor expliquen el rendiment de l'Oli són:

Variable	Coefficient
pH_inicial	0.0688
Gran1.5	0.04796
pH_14	-0.1011
pH_21	0.03407
Luteina	-0.01726
Zeaxantina	-0.02745
Epóxidos	-0.01912

Amb la taula de coeficients podem interpretar com afecten els canvis unitaris de les variables explicatives en el rendiment dels Olis. Fem una petita explicació a continuació:

- **pH\_inicial:** amb un coeficient positiu això implica que per cada increment en la variable resposta es veurà incrementada en un 0.069 en el rendiment.
- **pH\_14:** és la primera variable amb coeficient negatiu, aquí ho interpretariem com que a majors puntuacions d'aquesta variable menors puntuacions en la variable resposta.
- **pH\_21:** aquesta variable també disposa d'un coeficient positiu per tant també té una relació positiva amb la variable resposta.

Una curiositat a remarcar és el fet que el ph en diferents moments temporals són importants. Amb això el que interpretem és que a igualtat de condicions amb unes puntuacions de partida si aquestes es veuen alterades això és significatiu per al rendiment; seguidament en un segon moment temporal, restablint les puntuacions de  $ph_{t+x}$  com a base, a igualtat de condicions si es produeix algun canvi torna a ser significatiu, i així fins als 21 dies.

- **Gran1.5 :** amb un coeficient positiu això implica que per cada increment la variable resposta augmentarà en un 0.048 el rendiment.
- **Luteina:** amb un coeficient negatiu cada increment en la variable Luteina suposarà un decrement de 0.017 en la variable resposta.
- **Zeaxantina:** la variable resposta disminuirà un 0.027 cada vegada que la variable Zeaxatina augmenti.
- **Epóxidos:** amb un coeficient negatiu cada increment de la variable suposarà un decrement de 0.019 el rendiment.

## 3.2 Antioxidants

En aquest segon experiment es busca investigar com estabilitzat el producte (rendiment avaluat en diferents moments). Fins ara l'empresa treballava amb un antioxidant (producte per posposar l'oxidació), que els funcionava molt bé però que ha estat retirat del mercat. Com que no existeix cap anàleg conegut igual de bo volen provar diferents antioxidants.

En aquest cas es vol sotmetre cada oli a diferents antioxidants, mantenint les condicions atmosfèriques (temperatura) el més estable possible, per tal d'identificar la millor combinació possible, és a dir aquella que permeti que l'oxidació es retardi el màxim.

### 3.2.1 Mostra

En el segon experiment el que volem és seleccionar un bon antioxidant (Ferré, J., & Rius, F. X. 2002). Per fer-ho primer seleccionem 5 olis, seleccionarem els 5 olis a partir de les dades obtingudes anteriorment. Veiem a continuació un conjunt de dades que ens representa el rendiment dels olis en diversos moments temporals, concretament 6. Mostrem una petita mostra de les dades amb 5 files:

Etiquetas de fila	DÃa 7	DÃa 14	DÃa 21	DÃa 28	DÃa 35	DÃa 42
2018101901 (19660)	0.7215	0.5609	0.409	0.3856	0	0
2018101901 (19690)	0.6751	0.5537	0.4655	0.3925	0	0
2018101901 (19787)	0.7134	0.5603	0.3919	0.363	0	0
2018102001 (19642)	0.6098	0.477	0.3905	0.3326	0.2687	0

És interessant veure com a partir del dia 35 ja comencen a aparèixer valors a 0, més endavant veurem com tractem aquestes variables amb tants 0. A continuació veiem gràficament el nivell de rendiment del dia 7, a partir d'aquest gràfic seleccionarem els 5 olis que s'utilitzaran per a realitzar aquest segon experiment.

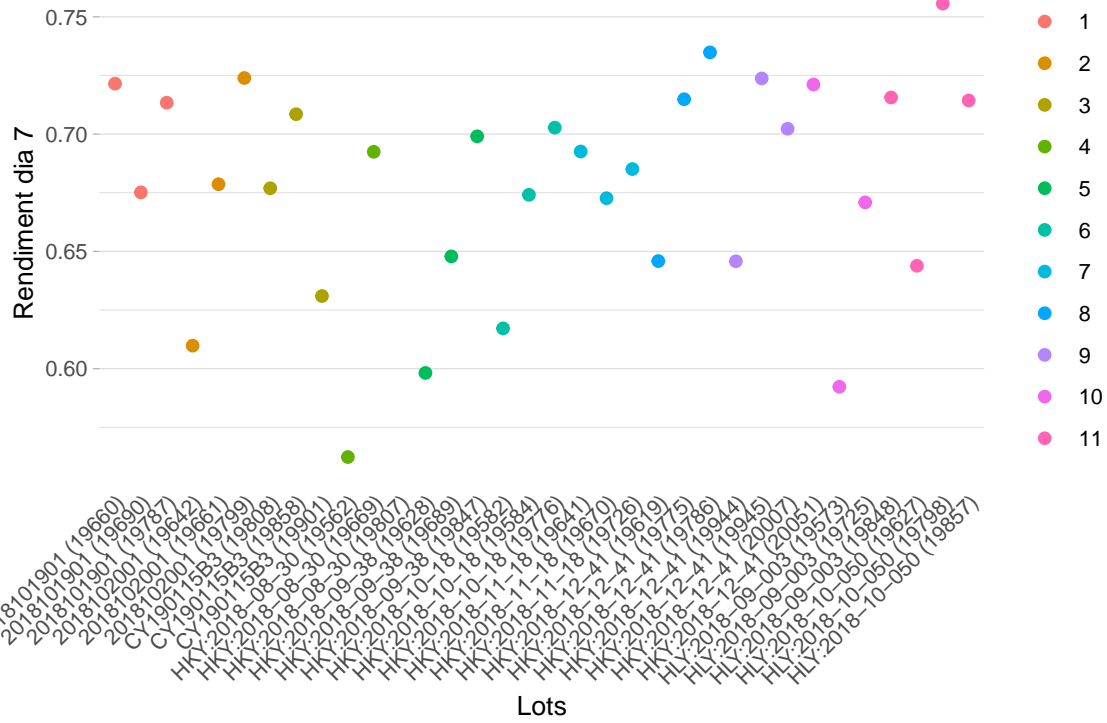


Figure 22: Selecció d'Olis

En aquest gràfic el que veiem són els orígens diferents amb 3 lots per cada un, d'entre aquestes tirades seleccionem els orígens amb observacions més interessants. Seleccionem un origen on els tres lots mostren rendiments molt diferents, un altre origen on els 3 rendiments donen resultats bastant elevats, el tercer i quart origen que seleccionem presenta rendiments mitjans tirant mentre que el cinquè origen mostra rendiments elevats en els tres lots.

Els identificadors i orígens amb els quals treballarem en aquest segon experiment són els següents:

Bidò	Origen
6298-1	Proveidor 1_1
6407-1	Proveidor 2_2
6311-2	Proveidor 1_2
6429-1	Proveidor 5
6487-2	Proveidor 3

Fem una representació gràfica dels rendiments que s'obtenen en la mostra seleccionada:

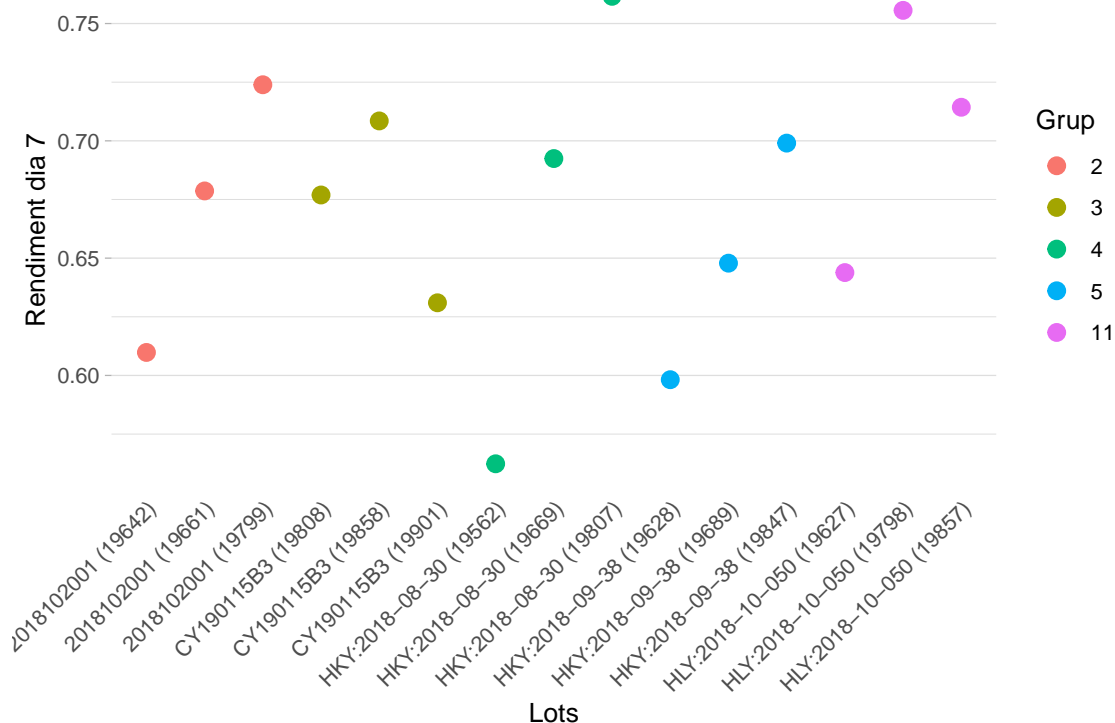


Figure 23: Seleció d'Olis

Dels 3 antioxidants seleccionats prèviament, el que farem serà diferents reaccions amb diferents antioxidants així com la categoria de control on no s'afegeix cap antioxidant.

El **procediment** és molt semblant al del primer experiment, de fet només es veu modificada una petita part:

1. Carregar el reactor amb el sac que toca.
2. Mesclar el producte i afegir l'antioxidant pertinent
3. Fer la reacció
4. Recollir la mostra i analitzar-ne la rendiment

### 3.2.2 Resultats

Intentarem avaluar les diferències en funció dels antioxidants, d'aquesta manera veurem si aquests ajuden a mantenir la qualitat del producte. A més a més estudiarem si hi ha diferències de comportament en funció dels olis.

#### 3.2.2.1 Anàlisis descriptives

Mostrem els resultats referents als antioxidants, ens interessa veure l'estabilitat del producte amb el temps en funció dels diferents antioxidants per veure si algun d'ells és millor. Mostrem una taula agregada per tipus d'antioxidant:

Aox	mean	N
AG 3%	0.6107	18
BHA 3%	0.7761	15
BHT:3%	0.7722	19
Blanc	0.6807	29

Amb aquesta taula semblaria que clarament BHT i BHA mantenen una millor qualitat, caldria veure si la diferència és estadísticament significativa i si es manté en el temps. El grup AG mostra una qualitat mitja inferior al grup de control.

#### 3.2.2.2 Anàlisis gràfiques

Fem una visualització de l'evolució del rendiment en funció de l'antioxidant:

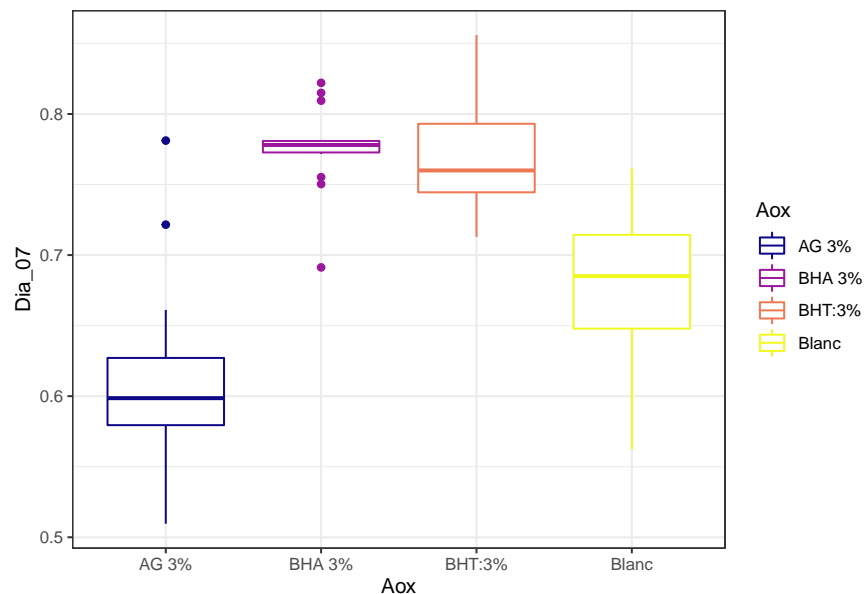


Figure 24: Del rendiment en funció dels Antioxidants

Gràficament semblaria que en el primer moment temporal registrat l'antioxidant BHA és el millor,

seguit per BHT, el grup control i per últim AG, a més a més en termes generals la retenció fluctua entre el 50-85%. Per veure com es comporten els antioxidants en el temps fem una gràfica on veiem els rendiments en un moment temporal avançat, i els valors de les puntuacions al cap de 35 dies:

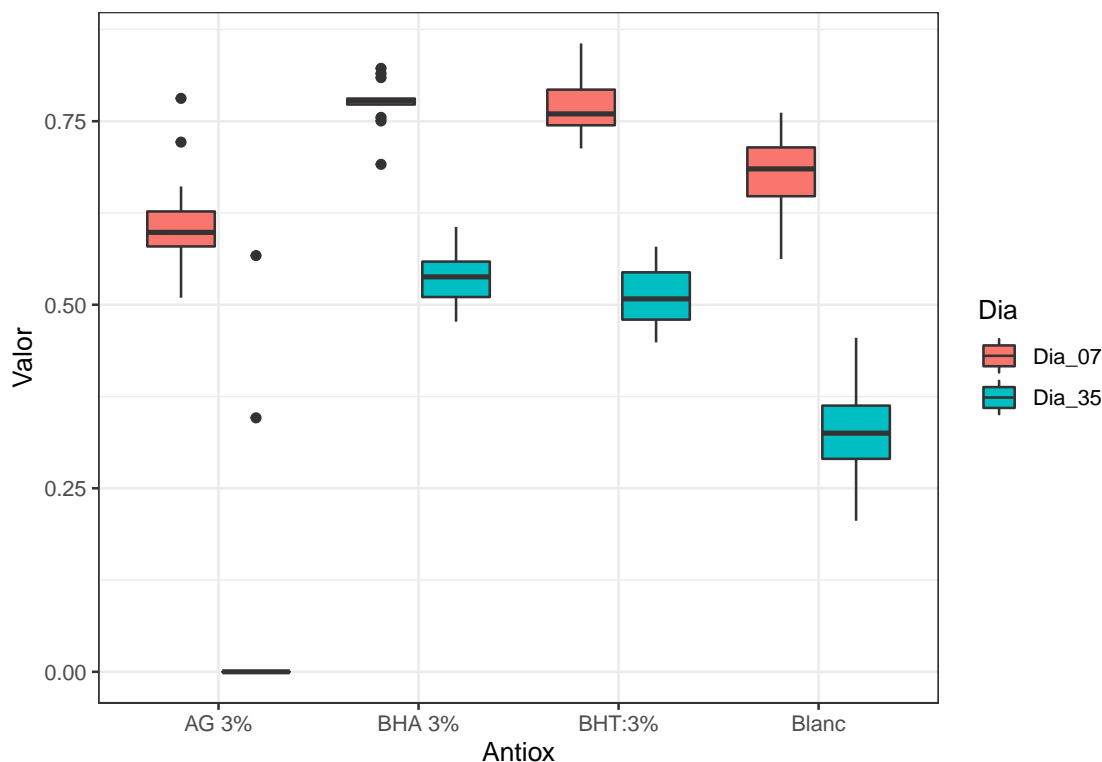


Figure 25: Retenció en el temps en funció dels Antioxidants

Al cap de 35 dies totes les mostres han empitjorat. Els olis que es veuen més afectats amb el pas dels dies són els del grup AG amb puntuacions de 0 en rendiment.

### 3.2.3 Modelització

Començarem amb una anàlisi de mesures repetides per veure si l'evolució al llarg del temps és diferent en funció de l'antioxidant. El que fem és ajustar un model de mesures repetides amb el temps i els reactius (Robbins, c.l 1985). Volem veure si a mesura que passa el temps hi ha diferències a la variable resposta en funció de l'antioxidant.

Per a realitzar de forma còmoda les anàlisis el que cal fer és construir 4 columnes corresponents a: codi individu, codi temps, antioxidants i variable resposta.

Ajustem un model de mesures repetides, per fer-ho utilitzem la funció `ezANOVA()` de la llibreria EZ. Aquesta funció requereix que indiquem com afecten els diferents factors, pel paràmetre *within* posem els diferents moments temporals i pel paràmetre *between* posem els antioxidants. Cal tenir present que aquesta funció requereix que els grups estiguin balancejats, i en aquest cas no ho estan. Presentem els resultats del anova:

```
ezANOVA(data = rr, dv = Valors, wid = ID, within = Data, between = Aox)
```

	Effect	DFn	DFd	F	p	p..05	ges
<b>2</b>	Aox	3	77	216.9	0	*	0.8185
<b>3</b>	Data	5	385	744.5	0	*	0.8185
<b>4</b>	Aox:Data	15	385	29.07	0	*	0.3457

Veiem que el factor Antioxidant, Data i la interacció són significatius. Aquesta conclusió no la podem considerar estrictament correcta, ja que les dades no són balancejades que és com la funció les requereix. A més a més per preservar la validesa del disseny cal confirmar que es compleix simetria composta per la taula de correlacions i les dades presenten Gynh-Feldt o esfericitat. Utilitzarem el test de Mauchly per avaluar l'esfericitat del model, cal tenir present que sovint no es compleix la condició amb aquest test, però ho provem:

	Effect	W	p	p..05
<b>3</b>	Data	0.2395	0	*
<b>4</b>	Aox:Data	0.2395	0	*

Amb el p-valor tan petit no tenim motius per no rebutjar  $H_0$  i per tant assumim que les diferències entre variàncies dels grups són diferents. Ergo s'està vulnerant la premissa d'esfericitat de l'estadístic F. En aquestes situacions el que hem de fer és corregir els p-valors obtinguts, concretament pel factor Data (*within*) i els de la interacció. El que fem és donar un cop d'ull als estadístics de Greenhouse-Geisser (el menys conservador) i Huynh-Feldt (IE, 016).

	Effect	GGe	p.GG.	p.GG...05	HFe	p.HF.	p.HF...05
<b>3</b>	Data	0.672	0	*	0.7063	0	*
<b>4</b>	Aox:Data	0.672	0	*	0.7063	0	*

Veiem que els p-valors pel factor Data i el de la interacció continuen sent significatius, així doncs basant-nos en aquests direm que són importants pel model i per tant que el model és bo.

Comparem els resultats amb una anàlisi del anova ajustant un model amb la funció `lme()`, la qual no requereix que les dades siguin balancejades per tant podrem comparar els p-valors, aquesta funció té un paràmetre específic per indicar el tipus de correlació que indiquem pel model, en aquest cas no especifiquem cap estructura:

```
lme(Valors ~ Data*Aox, random=~1|ID, data = rr, correlation = corSymm())
```

	numDF	denDF	F-value	p-value
<b>(Intercept)</b>	1	385	6503	0
<b>Data</b>	5	385	1008	0
<b>Aox</b>	3	77	163.8	0
<b>**Data:Aox**</b>	15	385	30.01	0

Amb aquesta funció alternativa obtenim que tant els factors com a la interacció entre ells és molt significativa.

Per veure si el model millora, utilitzem un cop més la funció `lme()` però aquesta vegada especifiquem una estructura AR1 de correlacions:

```
lme(Valors ~ Data*Aox, random=~1|ID, data = rr, correlation = corAR1())
```

	numDF	denDF	F-value	p-value
<b>(Intercept)</b>	1	385	7990	0
<b>Data</b>	5	385	604.1	0
<b>Aox</b>	3	77	217.9	0
<b>**Data:Aox**</b>	15	385	23.64	0

En aquest model també ens surten significatius els factors i la interacció.

Així doncs tenim dos models amb estructures de correlacions diferents, per comparar-los hem de reajustar els models amb el mètode de màxima versemblança en lloc de màxima versemblança restringida ( que és el que utilitza la funció `lme()` per defecte). Un cop aplicada la correcció podem comparar els models amb el AIC i BIC ajustat com hem fet en el primer experiment:

Variable	AIC	BIC
Model 1	-1426	-1254
Model AR(1)	-1383	-1254

En aquesta taula veiem que ambdós models tenen el mateix BIC, per tant el criteri queda descartat i és basem únicament en el AIC. Podem veure clarament que el model sense estructura de correlacions específiques presenta un AIC menor per tant considerem que és el millor model.

Fem una representació gràfica per veure com és comporta'n els factors i la interacció:

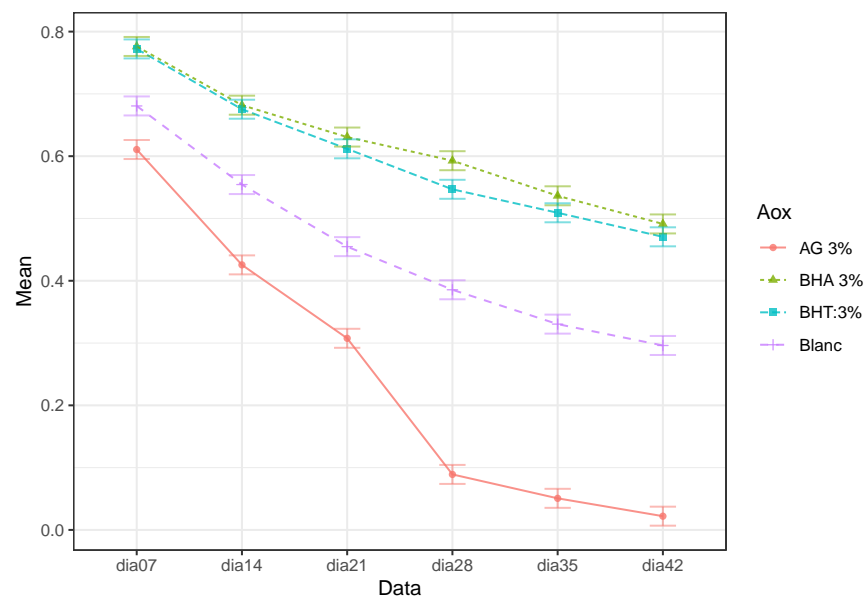


Figure 26: Anova Antioxidants



Es pot veure clarament en el gràfic que el grup AG no només és inferior des de l'inici a la resta de grups, sinó que a més a més disminueix molt ràpidament al llarg del temps. Pel que fa als altres grups podem veure que tant el BHA com el BHT mostren millors resultats que el grup de control (blanc). Com que el grup AG dona resultats molt inferiors el que fem és tornar a calcular l'Anova treient aquest grup del factor Antioxidant. Igual que hem fet prèviament, utilitzem la funció `lme()` sense estructura de correlació:

	numDF	denDF	F-value	p-value
<b>(Intercept)</b>	1	385	6503	0
<b>Data</b>	5	385	1008	0
<b>Aox</b>	3	77	163.8	0
<b>**Data:Aox**</b>	15	385	30.01	0

Amb aquests resultats veiem que tots els factors inclosa la interacció són significatius. Veiem a continuació com es comporten ara els factors:

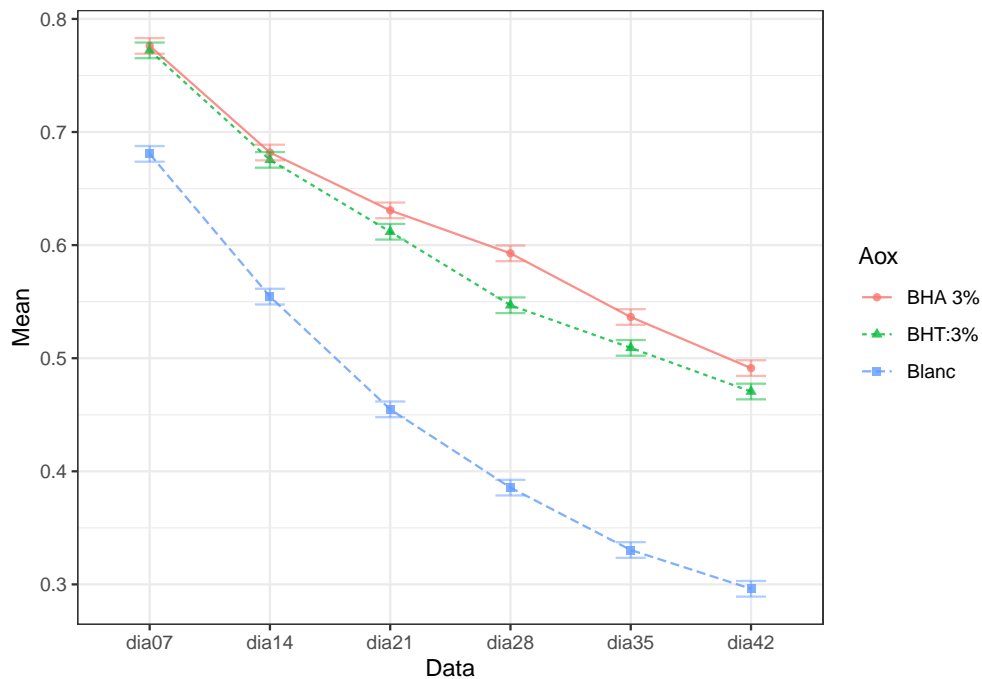


Figure 27: Anova redució nombre d'antioxidants

Com es veia en l'anàlisi prèvia veiem que tant BHT com BHA es comporten semblants, en els primers moments temporals mostren un comportament que sembla paral·lel i a més a més ambdós grup mostren valors molt semblants, a partir del dia 14 les puntuacions es distància una mica. El grup de control es manté des de l'inici amb resultats inferiors i a més a més el decreixement és més ràpidament. El següent pas, doncs, és determinar si les diferències entre BHT i BHA es poden considerar estadísticament significatives, per fer-ho el que farem és un nou model de mesures repetides amb els grups BHT i BHA:

	numDF	denDF	F-value	p-value
<b>(Intercept)</b>	1	160	15056	0
<b>Data</b>	5	160	503.2	0
<b>Aox</b>	1	32	2.078	0.1591
<b>**Data:Aox**</b>	5	160	7.395	0.000002898

En aquest contrast veiem que efectivament el factor antioxidant deixa de ser significatiu i per tant la interacció també. L'únic factor significatiu ara és el temps, en el gràfic ja podíem veure que en els diferents moments temporals la puntuació viaria bastant. Per tant la conclusió que trèiem és que ambdós antioxidants són iguals de bons. I per tant, la tria d'un o altre dependrà de qüestions de cost o disponibilitat.

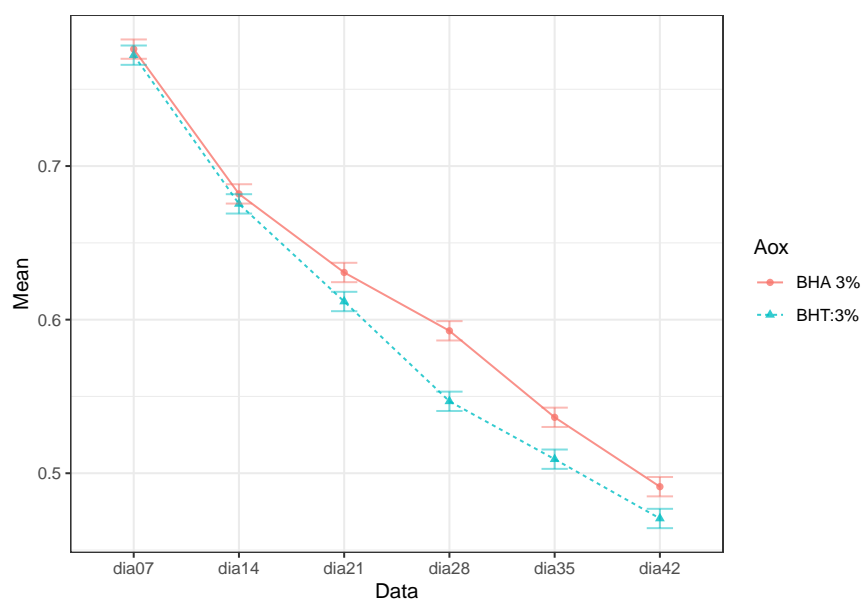


Figure 28: Anova BHT BHA

## 4 Conclusions

En ambdós experiments hem obtingut resultats interessants i amb repercussions directes a tenir en compte en l'operativa de l'empresa.

En la primera part el que volíem era identificar les característiques principals que identifiquen els millors olis. En aquest cas, a través d'un model, identifiquem quines són les variables explicatives rellevants per predir el rendiment així com el seu impacte sobre la variable resposta. Identifiquem les següents amb relació positiva: pH\_inicial, pH\_14 i Gran 1.5 això significa que augments en aquestes variables comporten puntuacions majors de rendiment. Per altra banda amb coeficients negatius trobem les variables pH\_21, Luteina, Zeaxantina i Epoxidos, en aquest cas puntuacions elevades comporten valors menors en la variable resposta. Així doncs si l'empresa realitza un petita anàlisi de les mostres que els ofereixen els proveïdors podrien seleccionar la matèria primera tenint en compte les puntuacions obtingudes en la variable resposta.

En el segon experiment hem comparat 3 antioxidants diferents i el grup control. En aquest experiment hem vist que tant la data com el tipus d'antioxidant que es fa servir té relació amb la retenció (estabilitat). Concretament hem vist que els antioxidants BHT i BHA són els millors, a més a més també veiem que l'Àcid Gàlic (AG) com a antioxidant és encara pitjor que el grup control on no s'afegeix res a la mescla. Un cop vist aquests resultats hem realitzat una petita prova per comprovar si hi havia diferències entre els dos millors antioxidants i hem vist que estadísticament es consideren iguals per tant la decisió d'utilitzar un o l'altre estarà basada en altres criteris com la disponibilitat o el preu.

## 5 Limitacions

Les principals limitacions han estat:

- Temps: a l'haver de combinar la disponibilitat de l'empresa amb els terminis del TFG, ha calgut ajustar o limitar els assajos per poder acabar a temps.
- Costos: el cost de les reaccions de cada prova també ha resultat un límit al nombre de mostres i repliques que hem pogut analitzar.
- Disponibilitat de l'empresa: tot i l'interès de l'empresa en el projecte, sovint hem trobat situacions imprevistes o altres prioritats que han alentit i limitat el projecte.

## 6 Recomanacions per futures investigacions

Pel que fa a l'experiment destinat a triar els millors olis, caldria utilitzar les conclusions de l'estudi per triar una sèrie de lots, repetir l'experiment i constatar si les prediccions sobre el rendiment resulten encertades o no (validació). Tanmateix fóra bo, poder analitzar altres característiques de l'oli i mirar si, amb aquestes, es pot millorar el model.

En el cas de l'experiment per triar antioxidant, caldria realitzar més proves per on es podria: afegir algun antioxidant més a la llista, provar combinacions entre productes, o provar diferents dosis d'inclusió de l'antioxidant.

## 7 Bibliografia

- Robbins, C. L. (1985). Daylighting. Design and analysis.
- Seber, G. A., & Lee, A. J. (2012). Linear regression analysis (Vol. 329). John Wiley & Sons.
- Kuehl, R. O. (2001). Diseño de experimentos: principios estadísticos para el diseño y análisis de investigaciones. Thomson Learning,.
- Ferré, J., & Rius, F. X. (2002). Introducción al diseño estadístico de experimentos. TECNICAS DE LABORATORIO-BARCELONA-, 648-653.
- Badiella, L. (2011). Modelos lineales generalizados mixtos: algunos casos prácticos. In Barcelona: X Congreso Galego de Estatística e Investigación de Operacións.
- Grant, T. (2019). Randomized Complete Block Design. PLANT-BREEDING-GENOMICS.
- IE. (016). Mauchly's Test of Sphericity in R. Tableau's Vision on Big Data.

### *Programacio en R*

- Cran-R. (2020). Retrieved from <https://cran.r-project.org/web/packages/ez/index.html>
- DataNovia. (2018). GGLOT EXAMPLES BEST REFERENCE. Retrieved from <https://www.datanovia.com/>
- RDocumentation. (2020). Retrieved from dataCamp Inc.